https://doi.org/10.5194/egusphere-2026-1960
Preprint. Discussion started: 20 May 2026


# GeoGen3D 1.0: An LMM-Based Reasoning Agent Framework for 3D Geological Model Generation

Jiateng Guo[1*], Junkun Li[1], Mark Jessell[2], Zhibin Liu[1], Luyuan Wang[1], and Xulei Wang[1]

[1]College of Resources and Civil Engineering, Northeastern University, Shenyang, China
[2]Centre for Exploration Targeting / Mineral Exploration Cooperative Research Centre / DARE Centre, School of Earth Sciences, The University of Western Australia, Perth, Australia

*Correspondence to*: Jiateng Guo (guojiateng@mail.neu.edu.cn; Tel.: +86-24-8368-7693)

**Abstract.** 3D Geological models provide conceptual and specific frameworks for a range of theoretical and applied geoscience activities, from theoretical research to the search for new resources. In many scenarios, the scarcity of data is common, and researchers must infer corresponding 3D geological structures based only on textual descriptions or outcrop images, for which standard 3D modelling approaches are poorly suited. Although current generative artificial intelligence can already generate pictures, videos, and 3D object models as required, there is still no feasible method to directly convert geologists' ideas or real photos of rock outcrops into 3D geological models. Here we present GeoGen3D, an intelligent Agent for text-image multimodal-driven 3D geological modeling. (1) Based on an improved ReAct agent framework, and by constructing a comprehensive collection of Noddy-based agent tools, we leverage the deep text and image understanding capabilities of large multimodal models (LMMs) to enable intelligent generation of 3D geological models from textual or visual inputs. (2) We introduce MMGM-Eval, a multimodal 3D geological model generation benchmark, to systematically evaluate the ability of LMMs to generate geological models from multimodal prompts. Our analyses demonstrate that GeoGen3D significantly outperforms direct prompt engineering approaches combining LMMs on the MMGM-Eval benchmark. GeoGen3D thus provides an efficient and intelligent modeling paradigm for multimodal-driven 3D geological model generation, especially suitable for scenarios lacking sufficient data.

## 1 Introduction

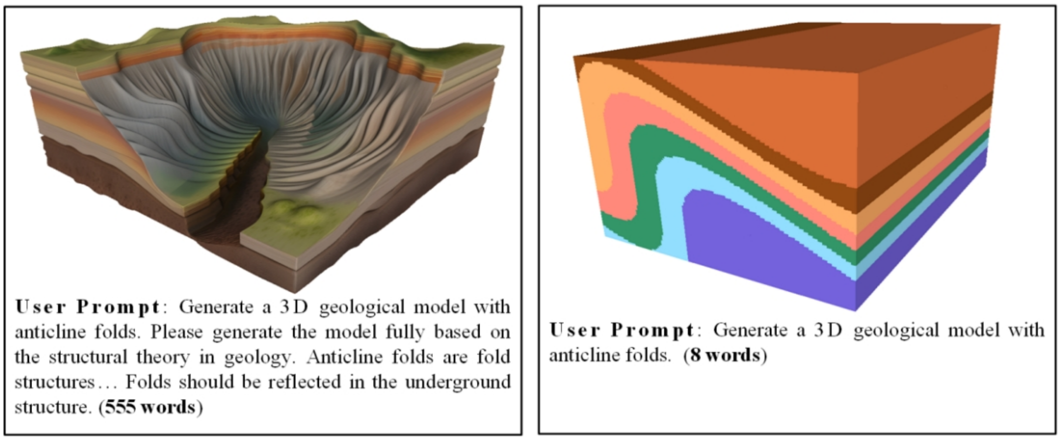
Three-dimensional geological models (Collon et al., 2015; Hassen et al., 2016) represent essential data products in modern Earth sciences, providing an indispensable digital framework for understanding and visualizing complex subsurface structures, stratigraphy, and resource distribution. These models integrate diverse geological data to construct spatial distribution representations of geological attributes such as stratigraphic units, lithology, and ore grade distribution. Three-dimensional geological models have extensive applications, including geotechnical investigation, engineering assessment, and scientific research.

Current mainstream three-dimensional geological modeling methods include explicit geological modeling methods, implicit surface geological modeling methods (Wellman and Caumon., 2018;de la Varga et al., 2019; Grose et al., 2021; M. J. Hillier

https://doi.org/10.5194/egusphere-2026-1960
Preprint. Discussion started: 20 May 2026


et al., 2014; Guo et al., 2021) and Deep learning geological modeling methods (Bi et al. 2022; M. Hillier et al. 2021; Yang et al. 2022; B. Zhang et al., 2023).

With the rapid development of Large Language Models (LLMs) and Large MultiModal Models (LMMs), artificial intelligence technology has created new approaches for solving complex scientific or engineering problems (OpenAI et al., 2024, 2025). Retrieval-Augmented Generation (RAG) technology (Karpukhin et al., 2020; Lewis et al., 2021) effectively mitigates the knowledge timeliness and hallucination problems of large models by integrating external knowledge bases and real-time information retrieval, enabling models to reason and generate based on the latest and most accurate information. AI Agent frameworks represented by ReAct (Yao et al., 2023) and Reflexion(Shinn et al., 2023) further expand the application boundaries of large models by endowing them with planning, reasoning, tool calling, and environmental interaction capabilities, enabling autonomous completion of complex multi-step tasks.




(a) Hunyuan3D Studio

(b) Ours

**Figure 1: Comparison with the generation effect of mainstream 3DAI models.**

While large-scale neural networks have achieved remarkable success in text-to-3D synthesis producing high-resolution, textured, and manifold representations driven by semantic prompts their application in geoscience remains challenging. Current models are predominantly trained on general-domain datasets, such as character assets and gaming props, and thus exhibit a significant deficit in understanding geological semantics, specifically the principles of tectonic geology. As depicted in Fig. 1, state-of-the-art generative models fail to capture the nuances of geological structures. Although the generated models possess exquisite textural quality, they are unable to correctly depict subsurface structural topologies. This domain gap leads to plausible-looking but structurally erroneous hallucinations, persisting even under rigorous semantic constraints.

To address the problem of automatically constructing three-dimensional geological models from geological semantic descriptions and real outcrop images, this paper proposes GeoGen3D, a three-dimensional geological modeling intelligent agent driven by multimodal large models. GeoGen3D utilizes an improved ReAct intelligent agent framework to iteratively reason about the formation of target geological structures, plan and execute a series of simulated geological structural events,

https://doi.org/10.5194/egusphere-2026-1960
Preprint. Discussion started: 20 May 2026


ultimately obtaining three-dimensional geological models consistent with textual descriptions or outcrop images. GeoGen3D can take user text instructions and field outcrop images as input, and through multiple structural evolution inferences, ultimately generate three-dimensional geological models with consistent structural morphology.

This study makes two core contributions:

(1) GeoGen3D Framework: We propose a GeoGen3D agent framework that can comprehensively understand multimodal inputs from textual descriptions and outcrop images, and autonomously generate corresponding three-dimensional geological models driven by geological events.

(2) MMGM-Eval Benchmark: We establish a novel Multimodal Geological Model Generation and Evaluation (MMGM-Eval) benchmark. This benchmark encompasses two categories of tasks: generating three-dimensional geological models from textual geological narratives and synthetic outcrop images, along with specific evaluation criteria and methodologies for task completion assessment.

The structure of this paper is organized as follows: Section 2 provides a detailed introduction to the overall architecture and specific methodologies of the GeoGen3D agent. Section 3 presents the MMGM-Eval benchmark constructed in this study and the corresponding test results of GeoGen3D, including two representative case studies.

## 2 Methods

### 2.1 GeoGen3D Structure

The problem addressed in this study can be described as follows: utilizing semantic geological structural text descriptions T or field outcrop images Ioutcrop as inputs to infer the set of geological activity events His that formed the structure, ultimately constructing the corresponding three-dimensional geological model M.

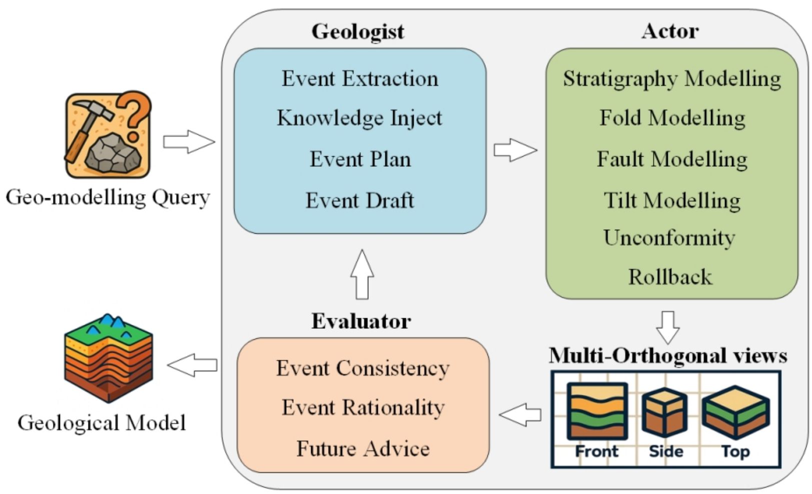



**Figure 2: Structure of GeoGen3D.**

https://doi.org/10.5194/egusphere-2026-1960
Preprint. Discussion started: 20 May 2026


The core approach of this work is to build upon a general-purpose multimodal large language model as the foundation, and implement targeted improvements to the ReAct Agent architecture based on specific geological modeling tasks, thereby developing a Geologist-Actor-Evaluator multi-agent collaborative workflow, as illustrated in Fig. 2. A geological event action space Anoddy is constructed based on the parametric geological modeling tool Noddy (Jessell, 1981; Jessell & Valenta, 1996). The Actor Agent implements intelligent selection of geological events, parameter generation, and modeling operations according to the parameter input rules of space Anoddy. The Evaluator Agent obtains the three-dimensional geological modeling results executed by the Actor in the previous step and provides evaluation of the previous Action's execution based on assessment criteria, which assists in decision-making for subsequent actions. Through the iterative process of Thought-Action-Observation, the system accomplishes complex geological structural modeling tasks that meet user requirements.

### 2.2 Geologist Agent Workflow

Within the methodological framework of this study, the Geologist Agent is primarily responsible for the following tasks: (a) Acting as a geologist, it must provide effective geological knowledge, modeling rules, and other relevant information to other agents in the framework based on user instructions before task initiation. (b) Prior to each modeling event conducted by the Actor, it must analyze and infer the next work plan according to the current state of the workflow, while simultaneously performing draft simulations of tectonic events under various parameter scenarios based on the planned content. To achieve these objectives, this study employs Prompt Engineering and Retrieval-Augmented Generation (RAG) methods to construct a comprehensive Geologist Workflow.

#### 2.2.1 Geo-Knowledge Injection

For the scientific task of generating three-dimensional geological models from multi-modal instructions in this specialized domain, directly employing mainstream general-purpose pre-trained multi-modal large language models, particularly those with smaller parameter counts, for inference and analysis may result in insufficient implicit prior knowledge in the geological domain, thereby causing hallucination effects in large models. Therefore, it is necessary to inject domain-specific knowledge related to user instructions into the Geologist Agent. This paper adopts a Geo-Knowledge Injection approach combining Wikipedia API with RAG, with the following workflow:

(a) **Query instruction generation**: Based on the rule descriptions of the overall process, the Geologist first responds to the user's multi-modal instructions by identifying keywords and questions that it believes need to be queried from external knowledge bases. The corresponding prompt for this step is shown in Fig. 3.

(b) **Knowledge document acquisition**: Using the keywords obtained from the previous step, encyclopedic knowledge is searched through the Wikipedia API to retrieve corresponding WIKI documents for the identified keywords.

(c) **Semantic retrieval of knowledge content**: The retrieved documents are rapidly segmented to obtain a collection of text fragments, and semantic retrieval is then employed to return text segments that are semantically most similar to the questions

https://doi.org/10.5194/egusphere-2026-1960
Preprint. Discussion started: 20 May 2026


posed in step (a). The specific process is as follows: First, all knowledge documents are segmented using a sliding window approach as shown in Eq. (1), where $S_i$ represents the $i$ text segment in the segmented text collection, $s$ denotes the sliding step size, and $l$ represents the segment length.

$$S_i = T[i \times s : i \times s + l], S_i \in S \tag{1}$$

$$D_i = BGE(S_i), D_i \in D \tag{2}$$

$$D_{TopK} = \underset{D_i \in D}{\operatorname{argmax}} \; {D_p}^T D_i \tag{3}$$

$$P_{init} = T + I_{outcrop} + P_{knowledgeQA} \tag{4}$$

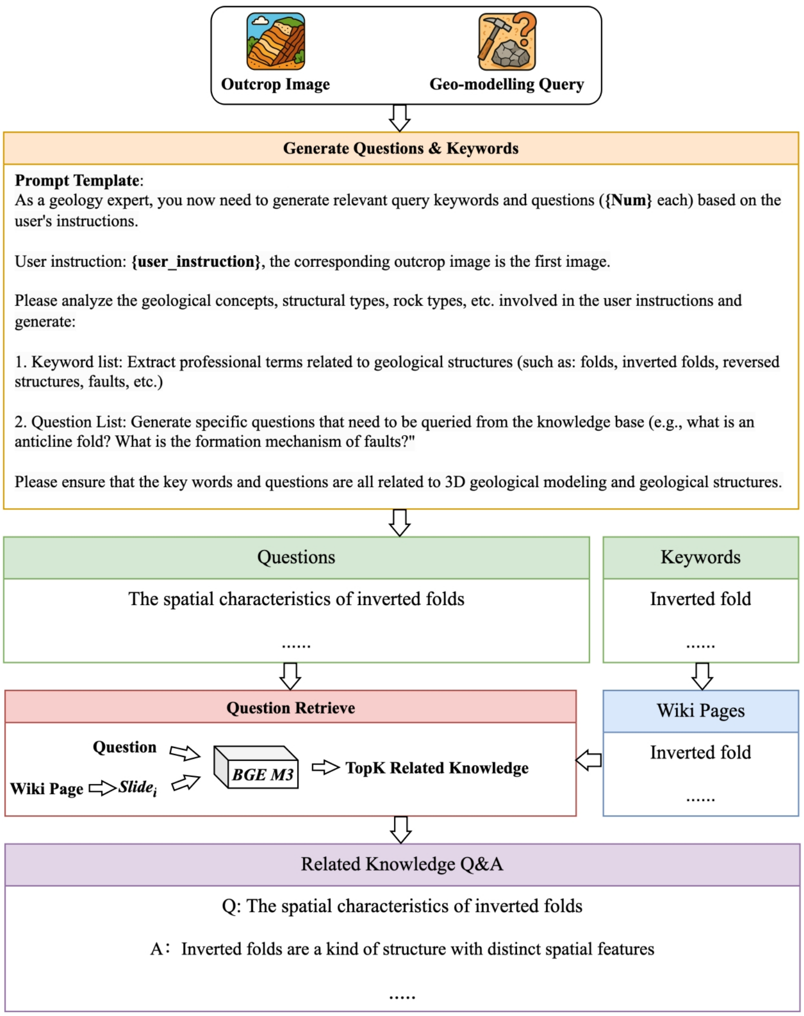



**Figure 3: Workflow of Geo-Knowledge Inject.**

https://doi.org/10.5194/egusphere-2026-1960
Preprint. Discussion started: 20 May 2026


To ensure semantic continuity during the segmentation process, this paper adopts $s < l$ to create overlapping character regions between segments. Next, the mainstream embedding model BGE M3(Chen et al., 2024) is employed to perform dense vectorization on each segment, constructing a segmented text vector collection $D$, as shown in Eq. (2). Finally, the questions obtained in step (a) are vectorized using the same method to obtain $D_p$. Maximum Inner Product Search (MIPS) is then applied to retrieve the top-K segments that are semantically most similar to the posed questions, as shown in Eq. (3). The retrieved relevant semantic information is integrated into $P_{knowledgeQA}$. Before the initiation of the overall workflow, the initial $P_{init}$ is constructed using the method shown in Eq. (4).

#### 2.2.2 Plan and Draft

Inspired by the chain-of-thought approach(Wei et al., 2023), before each modeling event undertaken by the Actor, the Geologist generates a corresponding Pplan based on the execution results of the previous step's event loop. This plan serves as reasoning and analytical support for the Actor to execute construction events throughout the workflow process.

In practical parametric geological modeling, relying solely on semantic modeling parameter rules as instruction inputs can make it challenging for Large Multimodal Models (LMMs) to truly comprehend the correspondence between individual parameters and actual modeling effects. Moreover, certain geological structural event modeling parameters, such as fold wavelength and fold amplitude, exhibit specific inter-parameter relationships that are difficult for LMMs to fully understand through textual descriptions alone, potentially leading to hallucinations. Therefore, we need to adopt a more concrete approach to enable LMMs to thoroughly understand the relationship between modeling parameters and modeling effects. Consequently, before the Actor executes each individual event, the Geologist provides Draft operation results based on the aforementioned Plan content. Specifically, the Draft process constructs a preliminary model according to the geological events and parameters predicted in the Plan. The draft model is built directly upon an exemplary sedimentary stratigraphic model, with geological structural events consistent with the Plan being directly added and geological evolution being performed using identical parameters. The six orthogonal views epresentation of the completed draft model is then provided by the Geologist as reference material for the Actor.

### 2.3 Actor Workflow

The Actor Agent serves as the core component within the GeoGen3D framework responsible for executing specific geological modeling tasks. Its primary function is to receive the plans and drafts generated by the Geologist Agent and transform them into precise instructions compatible with the Noddy parametric geological modeling engine. This workflow ensures seamless integration between high-level semantic planning and low-level parametric execution.

https://doi.org/10.5194/egusphere-2026-1960
Preprint. Discussion started: 20 May 2026


**Table 1**. Global Modeling Rules.

| Rule Category | Prompt Content |
|---|---|
| Coordinate System | X-axis: 0-10000 (from West to East direction);<br>Y-axis: 0-7000 (from South to North direction);<br>Z-axis: 0-5000 (from Downward to Upward direction); |
| Event Sequence | Stratigraphy must be executed first;<br>Stratigraphy can only be used once;<br>Complex structures result from event sequences (e.g., deposition → folding → faulting) |
| Parameter Format | All Action inputs must strictly follow specified JSON formats;<br>Use rollback when parameters don't meet requirements |
| Repetition Control | If similar action repeats >XX times unsuccessfully, consider alternative actions;<br>Maximum XX actions before forced termination |
| Physical Constraints | All geological events must respect physical feasibility;<br>Parameters should reflect realistic geological processes |

### 2.3.1 Actor Prompt

To effectively convey the reasoning results from the Geologist Agent to the Actor Agent and guide it in generating accurate Noddy modeling instructions, this paper designs a corresponding Actor Prompt. This prompt integrates multiple geological structural event sets supported by PyNoddy, a python wrapper for Noddy (Wellmann et al., 2016). Considering the LMM's comprehension ability, in order to slightly reduce the complexity of the overall process, we only applied some basic Noddy construction events in the Actor and did not cover all Noddy construction events. The composition of the Actor Prompt is shown in Eq. (5) and consists of the following components:

(a) The user's original multimodal instructions (text $T$ and/or image $I_{outcrop}$), ensuring that the Actor consistently operates toward the ultimate objective;

(b) The current step planning $P_{plan}$ generated by the Geologist Agent, providing the Actor with clear operational guidelines;

(c) The draft model view $V_{draft}$ generated through rehearsal by the Geologist Agent, offering the Actor a concrete parameter-effect reference that effectively reduces model hallucinations in parameter comprehension;

https://doi.org/10.5194/egusphere-2026-1960
Preprint. Discussion started: 20 May 2026


(d) The modeling rules $P_{rule}$ for structural geological evolution, where the most critical component $P_{rule}$ comprises the global modeling rules shown in Table 1.

$$P_{Actor} = T + I_{outcrop} + P_{plan} + P_{draft} + V_{draft} + P_{rule} \tag{5}$$

## 2.4 Evaluator Workflow

To ensure that models generated by GeoGen3D remain consistent with user intentions and to endow the intelligent agent with self-correction capabilities, we designed the Evaluator Workflow, as illustrated in Fig. 4. This workflow is activated after each geological event execution by the Actor, with its core function being to assess the effectiveness of executed actions and provide critical feedback information (Observation) for the next Thought-Action cycle.

After the Actor executes an Action, the framework invokes Noddy to compute and generate a new three-dimensional geological model. However, multimodal large models cannot directly process three-dimensional geological body model data. To enable the model to emulate human "vision" and comprehend 3D modeling results, we first render the generated three-dimensional model into a set of standardized two-dimensional views. Specifically, we generate six orthogonal views (front, back, left, right, top, and bottom views) along the positive and negative directions of the model's three principal axes (X, Y, Z). This set of views provides a comprehensive representation of the model's geometric morphology and stratigraphic contact relationships from all directions.

Subsequently, the Evaluator Agent is employed to assess the Action from the previous stage. We constructed an Evaluator Prompt that contains the following components: (1) the user's original multimodal instructions; (2) the complete Thought-Action historical records; (3) the six most recently generated model views. This prompt explicitly instructs the LMMs to serve as an Evaluator and make judgments primarily across the following dimensions:

(a) **Consistency assessment**: Whether the structural morphology displayed by the current model aligns with the user's textual descriptions, image features, or the Plan provided by the Geologist.

(b) **Correctness assessment**: The rationality of the Action executed in the previous step and its parameter selection, and whether it advances the model toward the final objective.

(c) **Future action advice**: Based on current assessment results, the actions that need to be undertaken in the next step.

The LMM's responses to the aforementioned questions are formatted as natural language text and provided as Observation feedback to the Geologist Agent. This Observation completes the Thought-Action-Observation loop within the GeoGen3D framework.

https://doi.org/10.5194/egusphere-2026-1960
Preprint. Discussion started: 20 May 2026

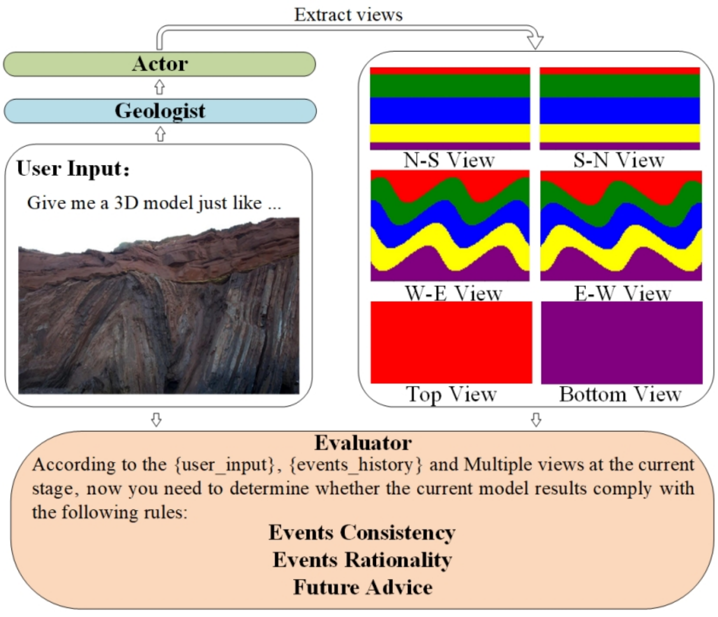




**Figure 4: Workflow of Evaluator.**

## 3 Experiments and Results

### 3.1 MMGM-Eval

To comprehensively evaluate the capabilities of Large Language Models (LLMs) and Large Multimodal Models (LMMs) in multimodal geological modeling tasks, this paper designs and constructs a specialized evaluation benchmark for instruction-driven three-dimensional geological model generation — MMGM-Eval (Multimodal Geological Model Generation Evaluation). This benchmark testing framework addresses the practical requirements for dual-modal inputs of textual descriptions and image information in geological scenarios, covers typical geological structure generation scenarios, and provides unified standards for evaluating the generalization capability and accuracy of different modeling approaches.

The MMGM-Eval benchmark comprises two major subsets: Text2Geo3D-Eval and Image2Geo3D-Eval, which systematically evaluate text-to-3D model generation and image-to-3D model generation capabilities, respectively. The specific organizational structure is as follows:

(a) **Text2Geo3D-Eval** subset contains 50 test cases. The textual content of each test case encompasses various geological structural conditions, including stratigraphic unconformities, faults, folds, and other geological formations, while specifying explicit requirements for geometric parameters (such as layer count, layer thickness, fault dip angle, etc.) and spatial proportional relationships. Each instruction is expressed in natural language form, focusing on examining multimodal large models' capabilities in understanding complex geological semantics, parameter parsing, and accurate generation of three-dimensional structural outcomes.

https://doi.org/10.5194/egusphere-2026-1960
Preprint. Discussion started: 20 May 2026


(b) **Image2Geo3D-Eval** subset comprises 25 carefully selected groups of real-world outcrop photographs from the internet, covering typical geological outcrop images from different regions, various shooting angles, and different structural complexities. These images encompass diverse geological types including complex folds, fault combinations, and different contact relationships, requiring models not only to intelligently extract and understand stratigraphic structures and structural morphological information from images, but also to accurately transform them into three-dimensional digital geological models.

To quantitatively evaluate the degree of matching between generated models and input instructions for each method, MMGM-Eval designs a standardized automatic scoring process. For each model generation task, we input the six principal views (front, back, left, right, top, bottom) of the generated three-dimensional model along with the corresponding multimodal input requirements directly into LMMs, enabling them to provide graded scores according to dimensions such as consistency, structural correctness, and numerical parameters. The scoring results are standardized to the [0, 1] interval, where 0 represents complete non-conformity, 1 represents complete conformity with geological descriptions or image features, and intermediate scores reflect partial matching or numerical approximation.

### 3.2 Results

To comprehensively evaluate the practical performance of GeoGen3D in multimodal 3D geological model generation tasks, this paper conducted systematic comparative experiments based on the MMGM-Eval benchmark between the proposed method and mainstream foundation large language models using Standard Prompt approaches (where all prompts are treated as a whole, requiring the LLM to directly output all geological event parameters sequentially in a single invocation to complete the modeling). Both approaches were tested under unsupervised zero-shot conditions. The evaluation covered both Text2Geo3D-Eval and Image2Geo3D-Eval subtasks, with experimental results presented in Table 2, 3, and 4, respectively. In each case the GeoGen3D process outperforms the DeepSeek, Gemini-2.5-pro and GPT-4o methods.

**Table 2**.Text2Geo3D-Eval Results.

| Method | Base Model | Score | | | Avg_Score |
|---|---|---|---|---|---|
| Standard Prompt | GPT-4o(OpenAI, 2024) | 24.25 | 21.75 | 25.50 | 23.17 |
| | GPT-4o-mini(OpenAI, 2024) | 15.25 | 14.50 | 15.25 | 15.00 |
| | DeepSeek V3(DeepSeek, 2024) | 22.50 | 22.50 | 23.00 | 22.67 |
| | Gemini-2.5-pro(Google, 2025) | 19.50 | 19.25 | 24.00 | 20.92 |
| **GeoGen3D** | GPT-4o(OpenAI, 2024) | 27.50 | 28.25 | 28.75 | **28.17** |

https://doi.org/10.5194/egusphere-2026-1960
Preprint. Discussion started: 20 May 2026


**Table 3**. Image2Geo3D-Eval Results

| Method | Base Model | Score | | | Avg_Score |
|---|---|---|---|---|---|
| Standard Prompt | GPT-4o(OpenAI, 2024) | 10.00 | 11.50 | 10.50 | 10.67 |
| | GPT-4o-mini(OpenAI, 2024) | 3.50 | 5.50 | 3.50 | 4.17 |
| | Gemini-2.5-pro(Google, 2025) | 10.00 | 10.00 | 9.50 | 9.83 |
| **GeoGen3D** | GPT-4o(OpenAI, 2024) | 17.00 | 16.50 | 16.00 | **16.50** |

**Table 4**. MMGM-Eval Total Results

| Method | Base Model | Text | Image | Total |
|---|---|---|---|---|
| Standard Prompt | GPT-4o(OpenAI, 2024) | 23.17 | 10.67 | 33.84 |
| | GPT-4o-mini(OpenAI, 2024) | 15.00 | 4.17 | 19.17 |
| | Gemini-2.5-pro(Google, 2025) | 20.92 | 9.83 | 30.75 |
| **GeoGen3D** | GPT-4o(OpenAI, 2024) | **28.17** | **16.50** | **44.67** |

Overall, GeoGen3D demonstrates substantial performance improvements across all evaluation metrics for multimodal geological modeling. These results validate the effectiveness of the proposed architecture in handling complex reasoning and generation tasks, thereby establishing a robust and viable framework for three-dimensional geological modeling in scenarios with limited direct exploration data.

### 3.3 Case Study

To more intuitively demonstrate the reasoning process and generation capabilities of GeoGen3D in handling complex geological modeling tasks, this section will introduce a representative case study in detail to demonstrate the internal automated reasoning process.

https://doi.org/10.5194/egusphere-2026-1960
Preprint. Discussion started: 20 May 2026

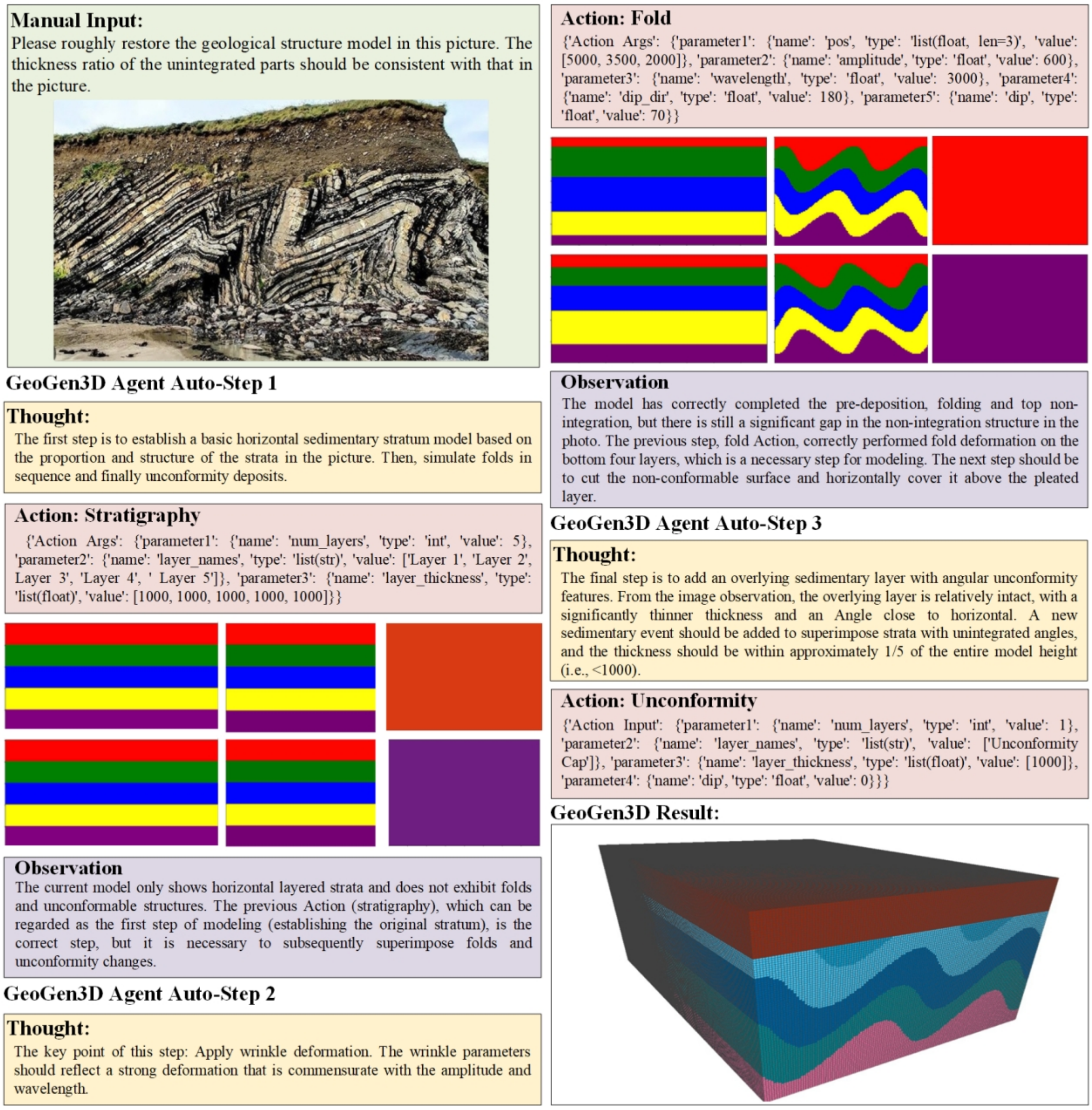




**Figure 5: Case of geological structure generation based on field outcrop images.**

This case study aims to evaluate GeoGen3D's multimodal understanding and geological reasoning capabilities when processing real-world visual data. The task input consists of a field outcrop photograph (upper left panel of Fig. 5.) that clearly displays intensely folded lower strata truncated by a near-horizontal erosional surface (unconformity), overlain by near-horizontal younger strata. As shown in Fig. 5., GeoGen3D successfully infers three-dimensional structures and their evolutionary history from two-dimensional imagery through multiple iterations:

Step 1: Construction of sedimentary strata. GeoGen3D begins with preliminary image analysis, with the Geologist determining that foundational strata should be established first. It explicitly indicates that the next step requires adding fold deformation.

https://doi.org/10.5194/egusphere-2026-1960
Preprint. Discussion started: 20 May 2026


Step 2: Construction of fold structures. Following the previous iteration's recommendations, the Geologist focuses on identifying and reproducing the fold morphology observed in the image, analyzing characteristics such as wavelength and amplitude. It invokes the Fold tool to generate the fold model. The subsequent Observation confirms the correctness of the fold geometry.

Step 3: Construction of unconformable strata. The Geologist enters the final geological reasoning phase, inferring that following fold formation, a process of structural uplift, erosion, and renewed deposition must have occurred, resulting in the formation of an unconformity surface. It invokes the Unconformity tool to superimpose an angular unconformity surface and overlying horizontal strata above the folded layers. The final Observation concludes that the generated model incorporates the three core structural features of folded basement, unconformity surface, and horizontal overlying strata, showing high consistency with the input image and completing the modeling task.

To better demonstrate the ability of this method to generate 3D geological models based on outcrop images, more practical examples are provided in Fig. 6.

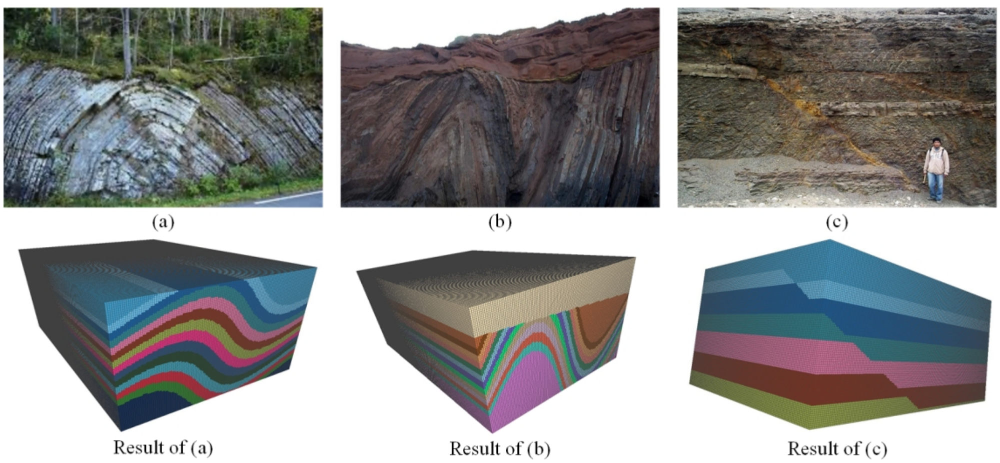



**Figure 6: Supplementary cases of 3D geological modeling for outcrop image generation.**

## 4 Discussion

This study developed the GeoGen3D multi-agent framework, which achieved significantly superior results compared to standard prompt methods across both task categories in the MMGM-Eval benchmark. These findings demonstrate that deep coupling of multimodal large language models with parametric geological modeling tools, enhanced by knowledge injection and iterative reasoning strategies, can effectively improve the parsing of complex geological semantics and model generation capabilities. Case studies further revealed that the Geologist-Actor-Evaluator iterative closed-loop enables GeoGen3D to dynamically adjust reasoning chains based on intermediate results, similar to human geologists. This approach provides a novel intelligent paradigm for mathematical geology methodologies.

https://doi.org/10.5194/egusphere-2026-1960
Preprint. Discussion started: 20 May 2026


The practical utility of GeoGen3D is manifested in two key aspects. First, geologists can leverage this framework during early research stages when high-density borehole or geophysical data are unavailable, utilizing textual records, historical atlases, or field photographs to rapidly construct three-dimensional conceptual models and conduct visual simulations. Second, GeoGen3D provides a highly semantic and customizable pipeline for three-dimensional geological model generation, capable of automatically producing large-scale datasets of 3D geological models, 2D geological maps, and virtual borehole profile data representing complex geological structural evolution.

Nevertheless, GeoGen3D has several limitations that require addressing in future work. First, although Noddy's event library covers common fundamental processes such as sedimentation, faulting, folding, and unconformities, its descriptive capacity is limited for higher-order or regional-scale structures such as salt domes and complex magmatic intrusions. In the case study, we can see that the predicted folds have a sinusoidal shape, unlike the chevron folds in the outcrop image, and the layer thickness varies in the output model, whereas the layers are essentially constant thickness in the image. PyNoddy could be altered to allow Noddy to have sharper hinge geometries, however the variations in layer thickness are beyond its capabilities. The use of a kinematic modelling engine such as Noddy has both strengths and weaknesses and there may be value in applying this approach using an implicit scheme such as LoopStructural which share's Noddy's ability to build time-aware 3D geological models. A broadly similar approach using RAG/LLM/Multi-agents is being developed using another modelling engine and in the future it will be informative to compare outcomes from these complementary approaches (Florian Wellmann, Pers. Com. 2025).

Second, the Evaluator's assessment criteria primarily rely on matching six-view projections with linguistic descriptions, lacking deep validation of quantitative geometric measurements, topological consistency, and physical evolutionary reasonableness, making it susceptible to large language model hallucinations and visual rendering ambiguities. A topological analysis of each model could in the future be added as an additional test of model quality. Additionally, the GeoGen3D workflow depends on the reasoning ability of multimodal models and the completeness of external knowledge bases; when domain knowledge entries are insufficient or image noise is excessive, agents may become trapped in ineffective iterations. Future research should consider incorporating physical knowledge constraints, expanding event sets and parameter spaces within more open structural grammars, and introducing hybrid geometric-topological evaluation metrics to further enhance the scientific authenticity and engineering applicability of the model. It is worth noting that GeoGen3D, as an Agent based on LMM, is limited by the temperature setting of LMM, its multimodal understanding ability, and the possible illusion effect of LMM. Therefore, for complex requirements, it maybe cannot achieve good modeling results in a single task. When the modeling results do not meet the requirements, adding more prompt words can achieve better results.

## 5 Conclusion

The GeoGen3D framework proposed in this paper deeply integrates the semantic understanding capabilities of multimodal large language models with the Noddy parametric geological forward modeling tool, establishing a "Geologist–Actor–

https://doi.org/10.5194/egusphere-2026-1960
Preprint. Discussion started: 20 May 2026


Evaluator" intelligent agent workflow that enables autonomous reasoning and generation from textual and visual inputs to three-dimensional geological models. Results from the MMGM-Eval benchmark demonstrate that GeoGen3D achieves significant improvements in average scores for both Text2Geo3D-Eval and Image2Geo3D-Eval tasks compared to large language models using standard prompts directly. Case studies validate that GeoGen3D can iteratively refine modeling parameters through multi-round reasoning processes, accurately reproducing typical structural evolutionary processes such as normal-reverse fault superposition and fold-unconformity combinations. This work provides a viable intelligent pathway for geological structural hypothesis validation and three-dimensional modeling based on multimodal information.

*Code and Data availability.* The source code of GeoGen3D method and the relevant experimental data in this research are available at: https://doi.org/10.5281/zenodo.19493634 (Guo and Li, 2026).

*Video supplement*. We have provided web links to download the video recordings of our case studies. The video supplement at: https://doi.org/10.5281/zenodo.19493634 (Guo and Li, 2026).

*Author contributions*. Jiateng Guo and Junkun Li conceived the manuscript; Jiateng Guo provided funding support and ideas; Junkun Li ; Mark Jessell and Zhibin Liu were responsible for the research method and program development; Mark Jessell, Luyuan Wang and Xulei Wang helped to improve the manuscript. All authors have read and agreed to the submitted version of the manuscript.

*Competing interests*. The authors declare that they have no conflict of interest.

*Acknowledgements*. This work was financially supported by the National Natural Science Foundation of China [grant number: 42172327]

https://doi.org/10.5194/egusphere-2026-1960
Preprint. Discussion started: 20 May 2026


## References

Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., and Anadkat, S.: Gpt-4 technical report, arXiv [preprint], arXiv:2303.08774, https://doi.org/10.48550/arXiv.2303.08774, 2023.

Bi, Z., Wu, X., Li, Z., Chang, D., and Yong, X.: DeepISMNet: Three-dimensional implicit structural modeling with convolutional neural network, Geoscientific Model Development, 2022, 1-28, https://doi.org/10.5194/gmd-15-6841-2022, 2022.

Chen, J., Xiao, S., Zhang, P., Luo, K., Lian, D., and Liu, Z.: M3-embedding: Multi-linguality, multi-functionality, multi-granularity text embeddings through self-knowledge distillation, Findings of the association for computational linguistics: ACL 2024, 2318-2335, https://doi.org/10.48550/arXiv.2402.03216, 2024.

Collon, P., Steckiewicz-Laurent, W., Pellerin, J., Laurent, G., Caumon, G., Reichart, G., and Vaute, L.: 3D geomodelling combining implicit surfaces and Voronoi-based remeshing: A case study in the Lorraine Coal Basin (France), Computers & Geosciences, 77, 29-43, https://doi.org/10.1016/j.cageo.2015.01.009, 2015.

de Kemp, E. A.: Spatial agents for geological surface modelling, Geoscientific Model Development, 14, 6661-6680, https://doi.org/10.5194/gmd-14-6661-2021, 2021.

De la Varga, M., Schaaf, A., and Wellmann, F.: GemPy 1.0: open-source stochastic geological modeling and inversion, Geoscientific Model Development, 12, 1-32, https://doi.org/10.5194/gmd-12-1-2019, 2019.

El-Kishky, A., Wei, A., Saraiva, A., Minaiev, B., Selsam, D., Dohan, D., Song, F., Lightman, H., Clavera, I., and Pachocki, J.: Competitive programming with large reasoning models, arXiv [preprint], arXiv:2502.06807, https://doi.org/10.48550/arXiv.2502.06807, 2025.

Grose, L., Ailleres, L., Laurent, G., and Jessell, M.: LoopStructural 1.0: time-aware geological modelling, Geoscientific Model Development, 14, 3915-3937, https://doi.org/10.5194/gmd-14-3915-2021, 2021.

Guo, J., Wang, X., Wang, J., Dai, X., Wu, L., Li, C., Li, F., Liu, S., and Jessell, M. W.: Three-dimensional geological modeling and spatial analysis from geotechnical borehole data using an implicit surface and marching tetrahedra algorithm, Engineering Geology, 284, 106047, https://doi.org/10.1016/j.enggeo.2021.106047, 2021b.

Guo, J. and Li, J.: LMMs can build 3D geological models: a reasoning agent framework of 3DGM generation, Zenodo [code, data set and video], https://doi.org/10.5281/zenodo.19493634, 2026.

Hassen, I., Gibson, H., Hamzaoui-Azaza, F., Negro, F., Rachid, K., and Bouhlila, R.: 3D geological modeling of the Kasserine Aquifer System, Central Tunisia: New insights into aquifer-geometry and interconnections for a better assessment of groundwater resources, Journal of Hydrology, 539, 223-236, https://doi.org/10.1016/j.jhydrol.2016.05.034, 2016.

Hillier, M., Wellmann, F., Brodaric, B., de Kemp, E., and Schetselaar, E.: Three-dimensional structural geological modeling using graph neural networks, Mathematical geosciences, 53, 1725-1749, https://doi.org/10.1007/s11004-021-09945-x, 2021.

Hillier, M. J., Schetselaar, E. M., de Kemp, E. A., and Perron, G.: Three-dimensional modelling of geological surfaces using generalized interpolation with radial basis functions, Mathematical Geosciences, 46, 931-953, https://doi.org/10.1007/s11004-014-9540-3, 2014.

Hong, Y., Zhen, H., Chen, P., Zheng, S., Du, Y., Chen, Z., and Gan, C.: 3d-llm: Injecting the 3d world into large language models, Advances in Neural Information Processing Systems, 36, 20482-20494, https://doi.org/10.48550/arXiv.2307.12981, 2023.

Hou, S., Zhao, A., Liang, J., Shen, Z., and Wu, H.: Geo-FuB: A method for constructing an Operator-Function knowledge base for geospatial code generation with large language models, Knowledge-Based Systems, 319, 113624, https://doi.org/10.1016/j.knosys.2025.113624, 2025a.

Hou, S., Shen, Z., Zhao, A., Liang, J., Gui, Z., Guan, X., Li, R., and Wu, H.: GeoCode-GPT: A large language model for geospatial code generation, International Journal of Applied Earth Observation and Geoinformation, 138, 104456, https://doi.org/10.1016/j.jag.2025.104456, 2025b.

Huang, Y., Gao, T., Xu, H., Zhao, Q., Song, Y., Gui, Z., Lv, T., Chen, H., Cui, L., and Li, S.: Peace: Empowering geologic map holistic understanding with mllms, Proceedings of the Computer Vision and Pattern Recognition Conference, 3899-3908, https://doi.org/10.48550/arXiv.2501.06184, 2025.

Jessell, M., Guo, J., Li, Y., Lindsay, M., Scalzo, R., Giraud, J., Pirot, G., Cripps, E., and Ogarko, V.: Into the Noddyverse: A massive data store of 3D geological models for Machine Learning & inversion applications, Earth System Science Data Discussions, 14, 1-19, https://doi.org/10.5194/essd-14-381-2022, 2022.

https://doi.org/10.5194/egusphere-2026-1960
Preprint. Discussion started: 20 May 2026


Jessell, M. W.: Noddy: an interactive map creation package, unpublished MSc thesis, University of London, 1981.

Jessell, M. W. and Valenta, R. K.: Structural geophysics: Integrated structural and geophysical modelling, in: Computer methods in the geosciences, Elsevier, 303-324, https://doi.org/10.1016/S1874-561X(96)80027-7, 1996.

Karpukhin, V., Oguz, B., Min, S., Lewis, P., Wu, L., Edunov, S., Chen, D., and Yih, W.-t.: Dense passage retrieval for open-domain question answering, Proceedings of the 2020 conference on empirical methods in natural language processing (EMNLP), 6769-6781, https://doi.org/10.48550/arXiv.2004.04906, 2020.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-t., and Rocktäschel, T.: Retrieval-augmented generation for knowledge-intensive nlp tasks, Advances in neural information processing systems, 33, 9459-9474, https://doi.org/10.48550/arXiv.2005.11401, 2020.

Lidal, E. M., Natali, M., Patel, D., Hauser, H., and Viola, I.: Geological storytelling, Computers & graphics, 37, 445-459, https://doi.org/10.1016/j.cag.2013.01.010, 2013.

Liu, H., Liu, J., Huang, S., Zhan, Y., Sun, H., Deng, W., Wei, F., and Zhang, Q.: se2: Sequential example selection for in-context learning, Findings of the Association for Computational Linguistics: ACL 2024, 5262-5284, https://doi.org/10.48550/arXiv.2402.13874, 2024.

Long, Q., Wu, Y., Wang, W., and Pan, S. J.: Does in-context learning really learn? rethinking how large language models respond and solve tasks via in-context learning, arXiv [preprint], arXiv:2404.07546, https://doi.org/10.48550/arXiv.2404.07546, 2024.

Natali, M., Viola, I., and Patel, D.: Rapid visualization of geological concepts, 2012 25th SIBGRApI conference on graphics, patterns and images, 150-157, https://doi.org/10.1109/SIBGRAPI.2012.29, 2012.

Shinn, N., Cassano, F., Gopinath, A., Narasimhan, K., and Yao, S.: Reflexion: Language agents with verbal reinforcement learning, Advances in neural information processing systems, 36, 8634-8652, https://doi.org/10.48550/arXiv.2303.11366, 2023.

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., Le, Q. V., and Zhou, D.: Chain-of-thought prompting elicits reasoning in large language models, Advances in neural information processing systems, 35, 24824-24837, https://doi.org/10.48550/arXiv.2201.11903, 2022.

Wellmann, J. F., Thiele, S. T., Lindsay, M. D., and Jessell, M. W.: pynoddy 1.0: an experimental platform for automated 3-D kinematic and potential field modelling, Geoscientific Model Development, 9, 1019-1035, https://doi.org/10.5194/gmd-9-1019-2016, 2016.

Wu, L.: Topological relations embodied in a generalized tri-prism (GTP) model for a 3D geoscience modeling system, Computers & Geosciences, 30, 405-418, https://doi.org/10.1016/j.cageo.2003.06.005, 2004.

Yang, Z., Chen, Q., Cui, Z., Liu, G., Dong, S., and Tian, Y.: Automatic reconstruction method of 3D geological models based on deep convolutional generative adversarial networks, Computational Geosciences, 26, 1135-1150, https://doi.org/10.1007/s10596-022-10152-8, 2022.

Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K. R., and Cao, Y.: React: Synergizing reasoning and acting in language models, The eleventh international conference on learning representations, https://doi.org/10.48550/arXiv.2210.03629, 2022.

Zhang, B., Xu, K., Khan, U., Li, X., Du, L., and Xu, Z.: A lightweight convolutional neural network with end-to-end learning for three-dimensional mineral prospectivity modeling: A case study of the Sanhetun Area, Heilongjiang Province, Northeastern China, Ore Geology Reviews, 163, 105788, https://doi.org/10.1016/j.oregeorev.2023.105788, 2023.

Zhang, Y., Wei, C., He, Z., and Yu, W.: GeoGPT: An assistant for understanding and processing geospatial tasks, International Journal of Applied Earth Observation and Geoinformation, 131, 103976, https://doi.org/10.1016/j.jag.2024.103976, 2024a.

Zhang, Y., He, Z., Li, J., Lin, J., Guan, Q., and Yu, W.: MapGPT: an autonomous framework for mapping by integrating large language model and cartographic tools, Cartography and Geographic Information Science, 51, 717-743, https://doi.org/10.1080/15230406.2024.2404868, 2024b.